63-4-5

DEPARTMENT OF PSYCHOLOGY

WASHINGTON UNIVERSITY

ST. LOUIS 5, MISSOURI

Partial Reduction Methods in Factor Analysis

FINAL REPORT

GRANT AF-AFOSR-62-181

30 JUNE 1963

This research was supported by the Behavioral Sciences Division, AFOSR, SRLB under ~~Contract~~/Grant AFOSR 62-181

STAFF:

Philip H. DuBois, Ph.D.
Principal Investigator

Research Psychologists
Robert F. Sullivan
Patricia Zwillinger

Research Assistant
Joseph S. Hupert

5 930 400

DEPARTMENT OF PSYCHOLOGY

WASHINGTON UNIVERSITY
ST. LOUIS, MISSOURI

Partial Reduction Methods in Factor Analysis

FINAL REPORT

GRANT AF-AFOSR-62-181

30 JUNE 1963

STAFF:

Philip H. DuBois, Ph.D.
Principal Investigator

Research Psychologists
Robert F. Sullivan
Patricia Zwillinger

Research Assistant
Joseph S. Hupert

Recapitulation

Contract AF 49(638)-614, between the Air Force Office of Scientific Research and Washington University, St. Louis, Missouri, became effective on 1 June 1959 with an original expiration date of 31 May 1960. The contract was subsequently extended to 31 December 1961. Scientific work was continued under Air Force Grant AF-AFOSR 62-181 from 1 January 1962 to 31 December 1962. This grant was subsequently extended to 30 June 1963.

The project title is "Partial Reduction Methods in Factor Analysis," with Philip H. DuBois, Professor of Psychology at Washington University, as Principal Investigator.

Project Staff

Under Contract AF 49(638)-614 research psychologists were Miss Elizabeth T. Ferrel, 1 June 1959-15 February 1960; Robert F. Sullivan, 1 July 1960-31 December 1961; and Mrs. Patricia Zwillinger, 1 July 1961-31 December 1961.

Under Air Force Grant AF-AFOSR 62-181 Mr. Sullivan was research psychologist from 1 January 1962 to 31 December 1962, and Mrs. Zwillinger was research psychologist from 1 January 1962 to 30 June 1962. In addition, Mr. Joseph S. Hupert served as graduate assistant during the closing phases in the spring of 1963.

Introduction

This project has been devoted to the study and development of the use of partial reduction techniques in factor analysis. These techniques are derived from the area of multivariate correlational analysis; their application to factor analysis yields certain procedures which have definite advantages over previous conventional methods. Partial reduction techniques are closely related to the diagonal method of factor analysis, a superior factor procedure because of its great logical simplicity.

Original objectives of this phase of the project included the following:

1. Precise specification of the characteristics of a correlation matrix and of the variables constituting the matrix for which the communalities are: a) determinable; and b) unique.

2. Definition of the minimum rank of a matrix in terms of the variables within the matrix and the structure of the underlying factor matrix.

3. Analysis of the relationship between minimum rank and the uniqueness and determinability of the communalities.

4. Delineation of the relationship between the minimum rank of a correlation matrix and the success with which partial reduction techniques can be applied to it.

5. Definition of the relationship between the uniqueness and the determinability of the communalities of a correlation matrix and the success with which partial reduction methods can be applied to it.

6. Determination of the sampling distributions of the statistics which were previously developed and specification of any required standard error formulae.

Research results have led to a concentration of effort on the first four of these objectives, since a satisfactory solution of the questions involved in these is requisite to any further analysis. Accordingly, attention has been focused upon two current and important topics in factor analysis, both relating to the problem of estimating communalities. First is an attempt to specify precisely the characteristics of a correlation matrix and of the variables constituting the matrix which admit of communalities that are determinable and/or unique. Second is the development of a refined empirical technique for the estimation of communalities.

Research Results

The concept of communality was first introduced by Thurstone in 1935. The logical elaboration of his rationale by several current factor theorists has led to the following definition and quantitative conclusions:

1. The communality may be defined as the squared multiple correlation of the given observed variables on the common factors.

2. The squared multiple correlation of the given variable on the remaining variables must be the lower bound to the communality if the Thurstone model is appropriate for the given matrix (Guttman, 1956, 1957).

3. Further, the communality is the upper limit of this squared multiple correlation as the number of variables approaches infinity, assuming constant rank (Guttman, 1957).

$$\lim_{j \to \infty} R_j^2 = h_j^2$$

4. Since the communality is a variance, its upper limit is unity, i.e.,

$$h_j^2 \leq 1.00$$

5. Unique communalities can be obtained only when the rank of the matrix satisfies the following inequality (Ledermann, 1937, Thurstone, 1947).

$$m \leq \frac{2n + 1 - \sqrt{8n + 1}}{2}$$

The above considerations provide a point of departure for both theoretical and empirical investigations.

The theoretical investigation of factors affecting the determinability and/or uniqueness of communalities was the subject of a Ph.D. dissertation by Mrs. Patricia Zwillinger. A copy of this dissertation will be submitted to AFOSR concurrent with presentation in fulfillment of academic requirements. Some results included here were taken from a draft of this dissertation.

A review of the literature indicated that the properties required of an acceptable set of communalities may be grouped in the following manner:

1. The obtained communalities must be within the following boundaries:

$$0 \leq R_j^2 \leq h_j^2 \leq 1.00$$

In empirical matrices the presence of error leads to the omission of equality signs, giving the equation

$$0 < R_j^2 < h_j^2 < 1.00$$

2. The factor matrix should produce the correlation matrix exactly:

$$R - U^2 = FF'$$

This condition implies invariance of the communalities regardless of the order in which factors are extracted. For empirical matrices the equation becomes

$$R - U^2 \sim FF'$$

This relationship, that of minimizing residuals, appears to be a common goal of factor analysts.

3. Minimum rank should be attained.

4. The correlation matrix with communalities in the diagonal must be Gramian.

An extensive analysis of a single empirical matrix was undertaken in order to compare the communalities produced by new and/or tested methods of approximation with those produced by more popular factor analytic techniques. The matrix chosen for extensive analysis was that produced by the MacQuarrie Test for Mechanical Ability. It was chosen on the basis of its previous presence in the literature, its simple definition and interpretation of its variables, and its convenient size (7 x 7).

The test consists of seven subtests: Tracing, Tapping, Dotting, Copying, Location, Blocks and Pursuit. The matrix of correlations among the seven MacQuarrie subtests is given in Table 1. These intercorrelations were first obtained by Goodman (1947) on a sample of 329 radio operators.

TABLE 1

Correlation Matrix for MacQuarrie Test for Mechanical Ability

Subtest	1	2	3	4	5	6	7
1 Tracing		.48	.55	.44	.34	.41	.43
2 Tapping	.48		.47	.31	.29	.29	.29
3 Dotting	.55	.47		.34	.43	.32	.36
4 Copying	.44	.31	.34		.54	.52	.48
5 Location	.34	.29	.43	.54		.54	.44
6 Blocks	.41	.29	.32	.52	.54		.46
7 Pursuit	.43	.29	.36	.48	.44	.46	

This correlation matrix was used for the purpose of illustrating various methods of obtaining communalities. The resulting sets of communalities were then compared and examined to see how adequately they met the aforementioned criteria for acceptability.

First, three "exact" solutions of the MacQuarrie matrix were described, two of rank six and one of rank five. A rank four solution was then found. A number of rank three solutions already available in the literature were examined and other, new rank three solutions calculated. One rank two and one rank one solution completed the survey. In each case, the factor matrix and communalities were tabulated. The correlation matrix was then produced from the factor matrix and residuals were calculated. A summary of these results is found in Table 2. Each set of communalities was put in the diagonal of the original correlation matrix and the inverse was taken.

TABLE 2

Comparative Summary of Obtained Communalities

Subtest	1	2	3	4	5	6	7
1 Tracing		.48	.55	.44	.34	.41	.43
2 Tapping	.48		.47	.31	.29	.29	.29
3 Dotting	.55	.47		.34	.43	.32	.36
4 Copying	.44	.31	.34		.54	.52	.48
5 Location	.34	.29	.43	.54		.54	.44
6 Blocks	.41	.29	.32	.52	.54		.46
7 Pursuit	.43	.29	.36	.48	.44	.46	
Multiple Correlation (R)	.670	.548	.642	.656	.664	.640	.589
Squared multiple R (R^2)	.449	.300	.412	.431	.441	.411	.347
Rank 6 Communalities							
DuBois-Sullivan I	.669	.497	.641	.667	.663	.637	.734
DuBois-Sullivan IIA	.674	.497	.641	.653	.663	.637	.488
Rank 5 Communalities							
DuBois-Sullivan IIB	.674	.497	.641	.536	.820	.506	.488
Rank 4 Communalities							
Modified Tryon-Kaiser A	.529	.861	.884	.540	1.000	.503	.437
Rank 3 Communalities							
Modified Tryon-Kaiser B	.512	.298	.884	.540	1.000	.502	.437
Centroid (Goodman)	.578	.469	.537	.548	.551	.534	.464
Principal Axes (Chapman)	.607	.420	.585	.550	.607	.526	.431
Canonical	.596	.414	.586	.547	.600	.530	.434
Albert	.679	.345	.521	.539	.976	.501	.434
Fortuitous	.831	.319	.991	.505	.586	.540	.407
4th Order Minor Expansion	1.074	.317	1.814	.555	2.187	.488	-.897
Rank 2 Communalities							
Spearman-DuBois	.583	.407	.543	.538	.570	.503	.428
Rank 1 Communalities							
Spearman Single Factor	.479	.283	.403	.469	.447	.430	.400

The following important results were noted:

1. Some of the communality estimates were extremely unstable (e.g., $.403 \leq h_j^2 \leq .991$). In this context, it was pointed out that when the rank of the matrix is greater than half its order, there always exist an infinite number of solutions. However, even when the criterion of $r \leq n/2$ was met, extreme variability was found (e.g., $.551 \leq h^2 \leq 1.000$). This variability could not be explained satisfactorily on the basis of error; rather it pointed out theoretical contradictions inherent in the various methods.

2. The relationship between rank and size of residuals proved to be an inverse one. That is, those solutions best meeting the minimization of residuals criterion tended to fall short on minimization of rank, and _vice versa_. This emphasizes the need for other criteria, such as Wrigley's (1961) _salience_.

3. The practical difficulties involved in obtaining a unique and determinable set of communalities were emphasized.

A theoretical analysis of the following question was then undertaken: Under what conditions may a correlation matrix be said to have _one and only one set_ of communalities? The relationship of two allied concepts, rank and complexity, to this question was also investigated.

Ledermann (1937) and Thurstone (1935) were the first writers to derive expressions defining those rank conditions under which communalities could be unique. Both authors arrive at the same formula although their lines of reasoning are somewhat different. This formula is also the same as that derived by Harman (1960). All three proofs involve the same assumptions and arrive at the same conclusions. An adaptation of Thurstone's presentation is given below.

An unique configuration of test vectors determines an unique set of communalities. In order to ascertain what rank conditions will permit an unique test configuration, it is necessary to know how many independent factors, m, can be precisely determined from any given correlation matrix of order n. To do this, the number of experimentally independent correlation coefficients among variables are compared to the number of linearly independent factor loadings in the factor matrix. The number of correlations in a correlation matrix is

$$\frac{n(n-1)}{2}$$

A configuration of test vectors can always be rotated so that the first axis is colinear with the first test vector, the second axis orthogonal to the first vector in the space spanned by the first two vectors, and so forth. This operation results in a number of zero factor loadings in the factor matrix. The number of such zero loadings is

$$\frac{m(m - 1)}{2},$$

and so the number of linearly independent factor loadings in a given factor matrix is

$$nm - \frac{m(m - 1)}{2}$$

A model correlation matrix and factor matrix are given in Table 3.

TABLE 3

Model Correlation and Factor Matrices

n x n Correlation Matrix R

1	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{17}
2		r_{23}	r_{24}	r_{25}	r_{26}	r_{27}
3			r_{34}	r_{35}	r_{36}	r_{37}
4				r_{45}	r_{46}	r_{47}
5					r_{56}	r_{57}
6						r_{67}
7						

n x m Factor Matrix F

	I	II	III
1	a_{1I}	0	0
2	a_{2I}	a_{2II}	0
3	a_{3I}	a_{3II}	a_{3III}
4	a_{4I}	a_{4II}	a_{4III}
5	a_{5I}	a_{5II}	a_{5III}
6	a_{6I}	a_{6II}	a_{6III}
7	a_{7I}	a_{7II}	a_{7III}

In order that there may be an unique configuration of test vectors for the correlation matrix, it is necessary for the number of experimentally independent constants (correlation coefficients) in R to equal or exceed the number of linearly independent parameters (factor loadings) in F. Hence,

$$nm - \frac{m(m - 1)}{2} \leq \frac{n(n - 1)}{2}$$

from which it can be seen that $\left[-m^2 + (2n + 1)m - n^2 + n\right] \leq 0.$

Solving for m,

$$\left(-m + \frac{2n + 1 + \sqrt{8n - 1}}{2}\right)\left(m - \frac{2n + 1 - \sqrt{8n - 1}}{2} \right) \leq 0$$

Assuming $0 \leq m < n$, the first expression above must be positive, therefore,

$$m \leq \frac{2n + 1 - \sqrt{8n + 1}}{2} \tag{1}$$

In other words, it is necessary that the above rank formula be satisfied in order for a correlation matrix of n tests to have an unique configuration in m dimensions. In a similar manner, it also follows from the original equation that the minimum number of tests required for the unambiguous determination of m factors is

$$n \geq \frac{(2m + 1) + \sqrt{8m + 1}}{2} \tag{2}$$

A careful examination of Thurstone's argument, given above, reveals certain problem areas. First of all, it is clear that the above relations are necessary for an unique configuration but not sufficient. They are required for but do not guarantee an unique configuration. Further specifications would be necessary before sufficient conditions could be defined. In addition, the stipulation is made that the correlations be <u>experimentally</u> independent. It will be shown in this section that the <u>linear</u> independence of the correlations is also an important factor related to the uniqueness of communalities. Most important of all, the restriction is made in this argument that the factor loadings must be linearly independent. However, the equations relating the factor matrix to the correlation matrix are of the following form:

$$a_{jI}a_{kI} + a_{jII}a_{kII} + a_{jIII}a_{kIII} = r_{jk}$$

These are not linear equations. Only Guttman (1958) seems to have recognized this important fact, and even he did not investigate its consequence. To be precise, these are quadratic equations; that is, they are second degree equations, not first degree equations. It would therefore appear more appropriate if the restriction be made that these equations have second degree independence, rather than linear independence.

Both Ledermann and Harman base their proofs on minor expansion solutions for communalities. Ledermann proves that minors of order (m + 1) are linearly independent, while Harman assumes that principal minors of order (m + 1) and (m + 2) are linearly independent. The principal minor expansions, however, do not yield linear equations, since the principal diagonal must always contain the unknown communalities. The first term in the principal minor expansion, then, is the product of (m + 1) or (m + 2) communalities, resulting in a (m + 1)th or (m + 2)th degree equation. For example, the expansion of the following 3 x 3 principal minor,

$$\begin{matrix} h_1^2 & r_{12} & r_{13} \\ r_{12} & h_2^2 & r_{23} \\ r_{13} & r_{23} & h_3^2 \end{matrix}$$

yields the following equation:

$$h_1^2 h_2^2 h_3^2 + 2r_{12} r_{13} r_{23} - r_{23}^2 h_1^2 - r_{13}^2 h_2^2 - r_{12}^2 h_3^2 = 0$$

This is a third degree equation in three unknowns, not a linear equation. Third degree independence is therefore required, rather than linear independence.

The type of independence required for the above proofs is not the only problem encountered in these attempts to define the relation between the rank of the correlation matrix and the uniqueness of the communality. Another question that can be raised is whether the assumption can be made that the rank of the whole correlation matrix is also the rank of various sub-matrices as well. An assumption of this kind will be shown to be necessary if the above solutions for the communality are to be determinate. However, a basic assumption of factor analysis is that the direct solution, resembling the factor matrix (in Table 3) used in Thurstone's proof above, can be rotated to a simpler structure. The underlying proposition involved here is that the complexity of most, if not all, variables is less than the rank of the correlation matrix (Thurstone, 1947). An implication of this proposition is that some of the factor loadings in the unrotated factor matrix are linearly dependent. Another implication is that any attempts to define the conditions under which communalities are unique must take into consideration not only the rank of the correlation matrix as a whole, but also the complexity of the variables and the factors involved as well.

When such complexity was investigated by Mrs. Zwillinger, certain inconsistencies in the development of commonly used criteria were noted. As mentioned earlier, the assumption cannot be made that all the variables, or more precisely, (n - m) variables, have complexity m, if the concept of simple structure is to be retained. This assumption appears to be made in the proofs described previously. It is made explicit in Thurstone's line of reasoning, but is not as obvious a part of the arguments of Ledermann and Harman. It can be shown, however, that the solutions these proofs are based on require that certain minors of order m be non-zero for all the communalities to be unique. This conclusion is based on the fact that matrices can be constructed which fulfill the necessary rank requirement Formula 1, but which still have variables with indeterminate communalities. In these cases, some but not all of the communalities were indeterminate, indicating that the defined rank requirements of Formula 1 actually guarantee only that unique communalities can be found for <u>some</u> of the variables in the matrix. For all the communalities to be unique, the factor complexity Formula 2 also has to be satisfied by sub-groups of factors. It can also be shown that unique communalities are not guaranteed even when the factor matrix satisfies both the rank requirements of Formula 1 as a whole as well as the factor complexity Formula 2 for all possible subgroups of factors.

In order that all the communalities be unique and determinate, there must be at least n(m x m) correlation submatrices containing linearly independent coefficients, one such matrix per variable, each located so that it contains none of the original diagonal entries (since these are unknown) and no correlations involving the associated variable. This stipulation applies only to matrices for which $r < n/2$. The minor expansion solution for the communalities must produce non-linear equations when

$$n/2 \leq m \leq \frac{2n + 1 - \sqrt{8n + 1}}{2}$$

In this situation, as n and m increase, the difficulties involved in finding unique solutions for the sets of simultaneous equations involved multiply rapidly.

A definition of the necessary and sufficient conditions for a set of unique communalities still was not attained. Such conditions, if they could be specified, would in essence allow a variable to have as few non-zero factor loadings as possible without permitting any factor or group of factors to be undetermined. It was shown that the well-known rank requirement formula:

$$m \leq \frac{2n + 1 - \sqrt{8n + 1}}{2} \qquad (1)$$

is a necessary but not a sufficient condition for unique communalities. This restriction does not allow for the varying complexity of variables and of factors which is possible if simple structure is to be assumed. In order to allow variables and factor loadings to have as few non-zero loadings as possible and still maintain unique communalities, further restrictions are required. An additional necessary condition was proposed, namely that the number of variables, n', with non-zero loadings on a group of m' factors, for all $m' \leq m$, be

$$n' \geq \frac{2m' + 1 + \sqrt{8m' + 1}}{2}, \; m' = 1,2,3, \ldots, m.$$

Each of these n' variables must have a non-zero loading on at least one (but not necessarily more than one) of the m' factors in the group. This condition has been informally recognized in the literature for $m' = 1$, but it does not seem to have been generalized beyond single factors.

Even these two necessary conditions taken together were shown not to be sufficient; multiple solutions are still possible for certain cases even when both conditions are met.

At this point, sufficient conditions can be proposed. It is contended that every associated correlation and factor matrix meeting the following three restrictions will have unique communalities:

1. For $(n \times n)$ R and $(n \times m)$ F, $m < \frac{n}{2}$

2. For F, $n' > 2m'$, $m' = 1,2, \ldots, m$, for all possible groups of m' factors in F.

3. For R, there must be at least $n(m + 1)(m + 1)$ non-zero minors, 1 per variable, containing no elements from the given vector and no diagonal elements from R.

These conditions are not necessary, however, since unique solutions can be constructed which violate them.

The difficulties encountered when

$$\frac{n}{2} \leq m \leq \frac{2n + 1 - \sqrt{8n + 1}}{2}$$

seem to permit no simple solution. All theoretical solutions for the communality become non-linear and may have multiple solutions. Unique and multiple solutions for matrices with the same order and rank which have this rank condition have been constructed. It appears that the non-linear equations defining the factor loadings in these cases are rather complicated, and the number of these equations and the number of non-zero minors necessary for unique communalities are a complex function of the amount by which the rank of the matrix exceeds $\frac{n-1}{2}$.

It would be well to point out that the fact that any of the statements regarding rank that have been made in this theoretical section will be applicable only to the factor matrices derived from real data, and will never be directly and precisely applicable to empirical matrices.

Concurrently with the theoretical investigation carried on by Mrs. Zwillinger, Mr. Robert Sullivan was refining a previously developed method of estimating communalities. This work is the subject of a Ph.D. dissertation, a copy of which will be submitted to AFOSR concurrent with presentation in fulfillment of academic requirements.

Historically the problem of determining suitable approximations of communalities arose with the advent of multiple factor analysis.

Harman (1960) approaches the general problem of the communality through the following considerations:

1. The conditions that the correlation coefficients must satisfy in order for their matrix to have a given rank;

2. The determination of communality under the assumption of the rank of the correlation matrix;

3. The theoretical solution for communality; and

4. The approximations to communality without prior knowledge of the rank.

He deals with these considerations in sequence by laying down the algebraic conditions for given ranks; theoretical and algebraic solutions of communalities assuming rank for artificial matrices, such as Albert's method; and of prime importance here, the iterative procedures of Kaiser, Guttman, Wrigley and others of approximating communalities without prior knowledge of the rank.

To demonstrate that there is no paucity of methods for estimating communalities a review of literature concerning communalities reveals the methods of estimating communalities listed in Table 4.

The intent here is not to add to this already long and varied list another method of estimating communalities, but to present a method of estimating communalities that has many unique characteristics.

As mentioned previously, the principal concern of conventional factor analysis is scientific parsimony. As DuBois states, "Most factor analysts are probably more interested in factoring a good approximation of the observed correlation matrix than they are in factoring the correlation matrix itself." One of the main concerns of this method is the observed correlation matrix and not the conventional factor generated matrix. Thus, the emphasis is on precision rather than rank reducing characteristics.

An important consideration, perhaps the most important, is that the estimated communalities obtained by the method to be

TABLE 4

1. Trial and Error Exact Formula. Assumption is that sampling error and rounding errors do not occur.

2. Exact Formula. Applies only to special cases. For example, a one factor solution.

3. Successive Approximations.

4. Co-linear Graphical Methods.

5. Centroid Number 1. Method which begins with the highest correlation.

6. Centroid Number 2. Method by which a test is projected on a group of tests having the highest correlation.

7. Three Test Formula. Uses a cluster of three highly correlated tests.

8. Four Test Formula. Expansion of the three test formula.

9. Summation Formula. Takes the variable that correlated the highest with a given variable.

10. Spearman Formulas.

11. Carlson Method. Employment of a graphical procedure.

12. Cayley Hamilton Theorem. An algebraic solution for communalities which states that any square matrix satisfies its own characteristic equation.

13. A. Albert's Method. A method of estimating communalities by the linear dependence of rows. It is based upon the assumption that the rank is less than half the order.

14. Multiple R and successive adjustments.

15. Tryon-Kaiser Solution for Communalities.

16. Modification of the Tryon-Kaiser Solution.

17. The Averoid Method.

18. Harris' Lower Bound.

described here are proper communalities. (Communality estimates are considered proper if they preserve the Gramian properties of the correlation matrix.) It can be demonstrated that other estimates of communalities do not preserve the Gramian properties of the observed correlation matrix.

Similar to the approaches of Guttman, Tryon and Kaiser, this method employs an iterative procedure of estimation, but is unlike them in that multiple R is used as the starting point instead of the Squared Multiple Correlation (SMC). Computationally it differs from these methods in that an analytical rearrangement procedure is used instead of the mathematically inconvenient method of solving for R or R^2 by obtaining the inverse of the correlation matrix.

The method developed utilized the following procedure:

1. With 1's in the principal diagonals, the zero order correlation matrix is reduced using pivotal condensation.

2. During the reduction phases the matrix is also rearranged at prescribed times in order that all of the multiple R's for all variables may be computed. (See Figure 1.)

3. During the iterative phases if any variance becomes negative, the amount by which it becomes negative is added to the variance in question and the complete reduction process is repeated.

4. After each complete reduction of all variables, the adjustment formula as listed below is employed for each variable.

$$V_i'' = \sqrt{V_i' V_j' - V_{ij}} \; *$$

where V_i' = Variance for a particular variable previous to reduction.

V_{ij} = The unpredicted variance for a variable after reduction.

V_i'' = The adjusted value to be substituted in the matrix for reduction.

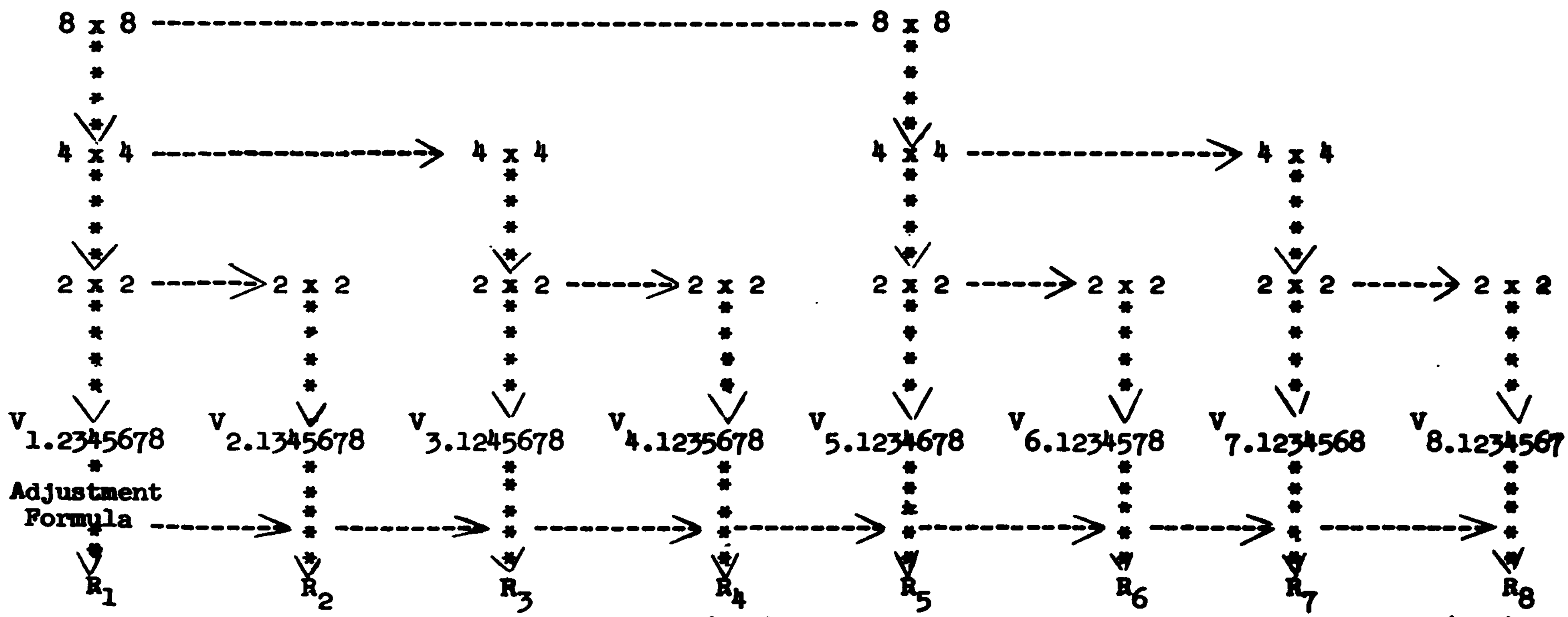
*It is to be noted that the first reduction and application of the adjustment formula would give multiple R.

5. The result of the above formula is then substituted into the diagonals of the zero order correlation matrix and the above steps are repeated until there is no change in the values of the diagonal elements. By no change it is meant that convergence is obtained either ideally or within a certain specified tolerance, e.g., .0005.

FIGURE 1

Analytic Rearrangement Procedure

Diagrammatically the Analytic Rearrangement Procedure is presented below. In this diagram the size of the matrix would have to be of the order of 8 or less. If it is less than 8, dummy variables would have to be employed (1.00's are placed in the diagonals and 0's in the off-diagonals.)



It is to be noted that all vertical lines (***) represent reductions and horizontal lines (---) represent rearrangements of the matrix.

From the diagram it is immediately evident that this procedure could be used for all sizes of matrices simply by increasing the number of reductions and rearrangements. Although the matrix may be of any size, it will be necessary to treat it as if it were of some power of 2. For example, a 9 x 9 matrix would involve a solution that would be capable of solving 16 x 16; a 17 x 17 would involve a solution that would be capable of solving a 32 x 32.

Pertinent results obtained by use of the formula include the following:

1. It was discovered that variation of the order of reduction of the matrix may result in either A) no convergence, and thus no solution, or B) alternate sets of exact communalities—that is, communalities which reproduce the original correlation matrix, leaving zero residuals. The alternate sets of communalities thus produced are not necessarily different for all variables. Invariance of h_j^2's despite varying order of reduction is obtained on the majority of variables. Usually, only one or two of the h_j^2's change.

2. One of the prime characteristics of the method of estimation described herein is that the matrix remains Gramian at all times. In order to demonstrate that other methods of estimation do not, in general, preserve the Gramian properties of the matrix throughout the entire estimation procedure, the MacQuarrie matrix was selected for further investigation. Communalities found by the following methods were examined.

A. Centroid
B. Principal Axes
C. Spearman
D. Albert
E. Modified Tryon-Kaiser
F. Single Factor
G. Canonical
H. Spearman-DuBois

The obtained h_j^2's were inserted as diagonal elements in the original correlation matrix which was then reduced. In all cases the matrix became non-Gramian.

3. Analyses of a variety of matrices were undertaken in an attempt to demonstrate the general applicability of the iterated and adjusted method. The following matrices were analyzed.

A. Guilford (1947) Aptitude Factor Analytic Study.
B. Spearman-Holzinger Unitary Trait Study (1934, 1935, 1936). (First eight variables.)
C. Artificial Matrix. (Thurstone, 1933.)
D. Mullen's Matrix (1939).
E. Guion Matrix (1953).

In all cases, the iterated and adjusted method produced convergence to exact solutions. In case A, alternate sets of exact communalities were produced. In case C, convergence could not be obtained with the variables in their natural order. Rearrangement of the variables, however permitted an exact solution of order N - 1. Case D was analysed in two distinct groups, producing identical solutions.

Attempts to use the method on matrices of higher order (e.g., 24 x 24) were unsuccessful, apparently because the singular characteristics of these matrices prevent maintenance of Gramian properties throughout the procedure.

Conclusions

One purpose of the present study has been to define those conditions under which a correlation matrix may be said to have one and only one set of communalities. Practical conditions discussed in the literature applicable to empirical data were often found to be contradictory (e.g., minimized rank versus minimized residuals), arbitrary (e.g., different specific criteria for minimized residuals), and, taken together, permitted a number of alternate sets of communalities, none of which completely satisfied all these criteria for acceptability. In addition, certain factorial methods were found to produce results which were not entirely as residual-minimizing as they were reputed to be. The ideal conditions applicable to artificial data were also found to be incomplete and to allow multiple solutions.

It appears that the conditions that are both necessary and sufficient for unique communalities still cannot be defined at the present time. This failure is in part due to the extreme complexity of the equations involved when the rank of the matrix is within the following range:

$$\frac{n}{2} \leq m \leq \frac{2n + 1 - \sqrt{8n + 1}}{2}$$

This may not be a serious fault, practically speaking, since the matrices found in empirical investigations generally are assumed to have a rank below these limits.

Several conclusions can be drawn as to the utility of the Thurstone factor model in view of the above results. First of all, the model is not entirely satisfactory from a theoretical point of view, since unique and determinate solutions cannot always be specified. The purpose of the model is to provide a rationale for the translation of a multitude of observed data to a modicum of basic dimensions, and this transformation must be logically precise and unambiguous; otherwise verification of the model is impossible. A certain amount of imprecision must be tolerated when dealing with empirical data, but the variety of distinctly different solutions which were found in the present study to be compatible with this model appears to be so great as to impair its utility seriously.

A second purpose has been to develop and refine a feasible method of estimating proper communalities. A method utilizing iteration and adjustment was developed which:

1. Was demonstrated to have a high degree of general applicability in producing exact solutions for communalities of order N - 1 and in some cases of order N - 2. These solutions are for empirical matrices.

2. When successful, accomplishes what no other method of estimating communalities does consistently; namely, preserves the Gramian properties of the original correlation matrix.

3. Meets the necessary conditions for determination of communalities.

4. Offers evidence for the existence of alternate exact sets of communalities.

Particular emphasis has been placed upon the desirability of analytic methods which preserve the Gramian properties of the matrix. It should be reiterated in support of this emphasis that the communality itself has only a theoretical existence; therefore, any method of estimating such a hypothetical construct must of necessity remain meaningful throughout. When a matrix becomes non-Gramian, the concept of negative variance is introduced and any results must be suspect. This problem does not exist when one is dealing with other than empirical matrices.

Documents produced in the course of the contract work are listed in Appendix A.

APPENDIX A

Documents Produced During the Course of the Contract

DuBois, P. H. On relationships between numbers and behavior. *Psychometrika*, 1962, 27, 323-333.

Marks, R. A. "A comparison of alternate methods of estimating communalities." M.A. thesis, Washington University, 1961.

Marks, R. A., and Zwillinger, P. G. "A modification of the Tryon-Kaiser method of finding communalities." Technical Report, Washington University, 1963.

Sullivan, R. F. "The use of proper communalities in factoring intercorrelations of psychological variables." To be submitted in partial fulfillment of requirements for the Ph.D. Degree, Washington University.

Zwillinger, P. G. "The determination of communalities in factor problems in psychology." To be submitted in partial fulfillment of requirements for the Ph.D. degree, Washington University.

BIBLIOGRAPHY

1. Albert, A. A. The matrices of factor analysis. Proc. Nat. Acad. Sci., 1944, 30, 90-95.

2. Chapman, R. L. The MacQuarrie test for mechanical ability. Psychometrika, 1948, 13, 175-180.

3. DuBois, P. H. Multivariate Correlational Analysis. New York, Harper and Bros., 1957.

4. DuBois, P. H. On relationships between numbers and behavior. Psychometrika, 1962, 27, 323-333.

5. Goodman, C. H. The MacQuarrie test for mechanical ability: II, factor analysis. J. Appl. Psychol., 1947, 31, 150-154.

6. Guilford, J. P., and Zimmerman, W. S. Some AAF findings concerning aptitude scores. Occupations, 1947, 26, 154-159.

7. Guion, R. M. Factorial classification of ability. Psychol. Bull., 1953, 50, 233-267.

8. Guttman, L. 'Best possible' systematic estimates of communalities. Psychometrika, 1956, 21, 273-286.

9. Guttman, L. Simple proofs of relations between the communality problem and multiple correlation. Psychometrika, 1957, 22, 147-157.

10. Guttman, L. To what extent can communalities reduce rank? Psychometrika, 1958, 23, 297-308.

11. Harman, H. H. Modern Factor Analysis, Chicago, University of Chicago Press, 1960.

12. Ledermann, W. On the rank of the reduced correlation matrix in multiple factor analysis. Psychometrika, 1937, 3, 85-93.

13. MacQuarrie, T. W. MacQuarrie Test for Mechanical Ability, Los Angeles, California Test Bureau.

14. Marks, R. A. "A comparison of alternate methods of estimating communalities." M.A. thesis, Washington University, 1961.

15. Marks, R. A., and Zwillinger, P. G. "A modification of the Tryon-Kaiser method of finding communalities." Technical report, St. Louis, Mo., Washington University, 1963.

16. Mullen, Frances. "Factors in the growth of girls seven to fourteen years of age." Ph.D. dissertation, Dept. of Education, University of Chicago, 1939.

17. __________, Preliminary report on Spearman-Holzinger unitary trait study, Nos. 1-9. Chicago Statistical Laboratory, Dept. of Education, University of Chicago, 1934, 1935, 1936.

18. Spearman, C. The Abilities of Man, New York, MacMillan Co., 1927.

19. Sullivan, R. F. "The use of proper communalities in factoring intercorrelations of psychological variables." To be submitted in partial fulfillment of requirements for the Ph.D. degree, Washington University.

20. Thurstone, L. L. The Theory of Multiple Factors, Ann Arbor, Edwards and Sons, 1933.

21. Thurstone, L. L. Vectors of the Mind, Chicago, University of Chicago Press, 1935.

22. Thurstone, L. L. Current misuse of the factorial methods, Psychometrika, 1937, 3, 73-76.

23. Thurstone, L. L. Multiple Factor Analysis, Chicago, University of Chicago Press, 1947.

24. Tryon, R. C. Communality of a variable: formulation by cluster analysis. Psychometrika, 1957, 22, 241-260.

25. Wrigley, C. Presentation at informal conference on "Necessary and Sufficient Conditions for Communalities," Washington University, St. Louis, Mo., December, 1961.

26. Zwillinger, P. G. "The determination of communalities in factor problems in psychology." To be submitted in partial fulfillment of requirements for the Ph.D. degree, Washington University.